

REMARKS

This application contains claims 1-35. Claims 4, 8, 19, 20, 25 and 32 have been canceled without prejudice. Claims 1, 5, 9, 11, 12, 15, 21, 22, 27, 29 and 34 are hereby amended. No new matter has been added. Reconsideration is respectfully requested.

Claims 1, 4, 8-13, 22, 25, 26, 29, 32 and 33 were rejected under 35 U.S.C. 102(e) over Halstead, Jr., et al. (U.S. Patent 5,963,893), while claims 2, 3, 23, 24,, 30 and 31 were rejected under 35 U.S.C. 103(a) over Halstead. Applicant has amended independent claims 1, 22 and 29 in order to clarify the distinction of the present invention over the cited art. Claim 1 has been amended to incorporate the limitations of claims 4 and 8, now canceled, while claims 22 and 29 have been amended in like manner. Claims 5, 9, 11 and 12 have been amended to depend from claim 1, following the cancellation of claims 4 and 8.

Claim 1, as amended, recites a method for disambiguation of an input string based on the morphological pattern of the string. The string is morphologically analyzed to generate a list of candidate analyses, each analysis comprising a word and its linguistic pattern. The pattern is defined in the claim as comprising at least one of the part of speech, prefix, number, gender and person of the word. The analyses are then selected based on the relative frequency of occurrence of their respective patterns. The technique permits each word to be analyzed and disambiguated individually, independently of its context. It may be used, *inter alia*, to prune the number of analyses of the words in a document that are included in a search index (page 6, lines 6-9, in the specification), so that only relatively frequent patterns are used.

Halstead describes a word breaking facility for identifying words within a Japanese text string based on

morphological processing (abstract). In one aspect of this facility, n-gram templates are provided that specify a pattern of character types found in stems and the likelihood of the template occurring in a stem (col. 1, line 64 - col. 2, line 1). Morphological analysis is performed on an input string of Japanese characters to give a directed acyclic graph. Each path through the graph represents a morphological analysis of at least a portion of the input string. The paths are scored to favor paths that include analysis for a greater portion of the input string and paths that include a greater number of bound morphemes. The highest-scoring path is selected and applied as the morphological analysis of the input string (col. 2, lines 13-27).

Thus, Halstead applies morphological analysis to an entire text string in order to find breaks between different words in the string. The analysis that Halstead ultimately selects for each word in the string is intimately connected with the analyses chosen for the other words in the string. One of the tools that he uses for this purpose involves the relative frequency of certain "ophemes," which are "character-type template patterns that are used during stem analysis to identify stems" (col. 3, lines 64-67). In other words, Halstead uses the frequency of occurrence of specific patterns of Japanese characters in word stems in order to determine which groups of characters in the input text string are likely to be stems (col. 10, lines 10-17, cited by the Examiner).

On the other hand, Halstead makes no use of the frequency of the morphological patterns of the individual words as an input to his analysis, as defined by amended claim 1. As noted above, the claimed patterns are explicitly defined as relating to specific linguistic characteristics: part of speech, prefix, number, gender and person. These characteristics are independent of and

indifferent to the stems (or lemmas) of the words in question. Halstead neither teaches nor suggests that the frequency of the morphological pattern, as defined in claim 1 (as opposed to the frequency of character patterns), might be used as a criterion in selecting an analysis of a word.

Thus, claim 1, as amended, is believed to be patentable over Halstead. In view of the patentability of claim 1, claims 2, 3 and 9-13, which depend from claim 1, are believed to be patentable, as well.

Claims 22 and 29 recite, respectively, a computer software product and apparatus that operate on principles similar to the method recited in claim 1. These claims were rejected on grounds identical to the grounds of rejection of claim 1, and have now been amended in like manner to claim 1. Therefore, Applicant respectfully submits that claims 22 and 29, as amended, are patentable over Halstead, as are claims 23, 24, 26, 30, 31 and 33, which depend from claims 22 and 29.

Claims 15, 16, 19, 27, 28, 34 and 35 were rejected under 35 U.S.C. 102(b) over Zamora (U.S. Patent 4,862,408), while claims 17 and 18 were rejected under 35 U.S.C. 103(a) over Zamora. Applicant has amended independent claims 15, 27 and 34 in order to clarify the distinction of the present invention over the cited art. Claim 15 has been amended to incorporate the limitations of claims 19 and 20, now canceled, while claims 27 and 34 have been amended in like manner. Claim 21 has been amended to depend from claim 15, following the cancellation of claim 20.

Claim 15, as amended, recites a method for searching a corpus of text. The method comprises morphologically analyzing the words in the corpus to generate a list of candidate analyses, and evaluating the pattern of each of the analyses in order to select one or more of the analyses from the list for each word. As in amended

claim 1, the pattern is defined here as comprising at least one of the part of speech, prefix, number, gender and person of the word, and the analyses are selected based on the relative frequency of occurrence of their respective patterns. The lemmas of the selected analyses are entered in an index of the corpus, to which a search query may then be applied.

Zamora describes a method for analyzing text using a paradigm. He creates a file structure in which each word in a list of words (or "dictionary") is associated with a set of paradigm references. These references generate all forms of each of the lemmas of the words in the list (abstract). In other words, Zamora uses all possible linguistic forms of each of the words in a given list (col. 2, lines 66-68, cited by the Examiner), without discriminating between the more and less frequent forms, as required by amended claim 15.

In rejecting claim 20, the Examiner maintained that Zamora discloses (in col. 7, lines 36-39) the limitation of determining a relative frequency of occurrence of the pattern of each of the analyses, as is now recited in amended claim 15. Applicant respectfully disagrees with this analysis. The passage in question relates to "front encoding," in which dictionary entries are compacted by counting how many leading characters in one entry are repeated in the next entry. This technique relates to the characters in the words, and has nothing to do with determining the frequency of linguistic patterns of the words, as defined in claim 15.

Thus, claim 15, as amended, is believed to be patentable over Zamora. In view of the patentability of claim 15, claims 16-18, which depend from claim 15, are believed to be patentable, as well.

Claims 27 and 34 recite, respectively, a computer software product and apparatus that operate on principles similar to the method recited in claim 15. These claims

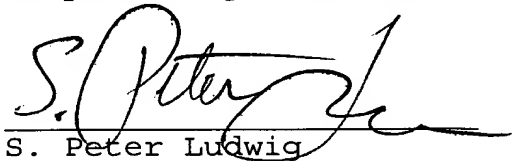
were rejected on grounds identical to the grounds of rejection of claim 15, and have now been amended in like manner to claim 15. Therefore, Applicant respectfully submits that claims 27 and 34, as amended, are patentable over Zamora, as are claims 28 and 35, which depend from claims 27 and 34.

Claims 5-7, 14, 20 and 21 were rejected under 35 U.S.C. 103(a) over the combination of Halstead and Zamora. Claim 20 has been canceled. In view of the patentability of claim 1 over Halstead and the analysis of the independent claims given above, claims 5-7, 14 and 21 are believed to be patentable over this combination of references.

Applicant has studied the additional references made of record by the Examiner, and believes the claims in this application to be patentable over these references, as well, whether they are taken individually or in any combination.

Applicant believes the amendments and remarks presented hereinabove to be fully responsive to all of the grounds of rejection raised by the Examiner. In view of these amendments and remarks, Applicant respectfully submits that all of the claims in the present application are in order for allowance. Notice to this effect is hereby requested.

Respectfully submitted,

A handwritten signature in dark ink, appearing to read "S. Peter Ludwig", is written over a horizontal line.

S. Peter Ludwig  
Registration No. 25,351  
Attorney for Applicant

Darby & Darby P.C.  
805 Third Avenue  
New York, NY 10022  
(212) 527-7700